Forthcoming in *Georgetown Journal of Law and Public Policy*

# Nudges and Bumps

Victor Kumar
University of Toronto

You probably know that the number of people waiting for organ transplants in the U.S. exceeds the number of organ donors. You may not know that the number of people who support organ donation also exceeds the number of actual donors (Kurtz and Saks 1996). Why, then, is there a shortfall in the pool of organ donors? And how can this large-scale social problem be fixed? A clue to answering both questions can be found in studies indicating that countries with an opt-out policy for organ donation have significantly higher rates of organ donation than counties with an opt-in policy (Johnson and Goldstein 2003; Abadie and Gay 2004). "Opt-in" means you have to sign a form to become an organ donor; "opt-out" means you have to sign a form *not* to become one. Evidently, people are more likely to stick with the default, whatever it happens to be.

Richard Thaler and Cass Sunstein (2008: 177-9) suggest that the U.S. might consider nudging would-be organ donors by switching from an opt-in policy to an opt-out policy (among other possibilities). People would be automatically assigned to organ donation unless they opted out, and many people would fail to opt out. So, this "nudge" would increase rates of organ donation, and it would do so without any coercion, since people can still refuse to be organ donors if they so wish. Not just morally desirable, switching to an opt-out policy would also seem to be highly effective—more effective than forcing people to be organ donors or creating economic incentives. On the one hand, requiring organ donation is coercive and invites political resistance. On the other hand, incentivizing organ donation is financially costly and perhaps even counterproductive. Offering cash for prospective organs might lead citizens to treat organ donation as an optional market transaction rather than an obligatory moral responsibility, further decreasing the number of willing organ donors. A nudge can do the job better.

In their work on nudges, Thaler and Sunstein crystallize an idea that promises to have important and far-reaching policy applications. A nudge, they explain, is an "aspect of choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (2008: 6). According to Thaler and Sunstein, governments, firms, and other "choice architects" can use nudges to promote well being but without infringing upon freedom of choice. A nudge, it seems, is a *moral loophole.*

Thaler and Sunstein do not provide an adequate definition of nudges, and I will begin the essay by developing one. Once we arrive at an adequate definition, it becomes clear that nudges have a significant and unacknowledged moral shortcoming: nudges do not straightforwardly infringe upon freedom of choice, but they nonetheless violate autonomy by circumventing rational agency. In a word, nudges are *manipulative*. Recent work in the science of judgment and decision-making, however, suggests an alternative. A *bump*, like a nudge, structures choice architecture so as to influence behavior, but it does so in ways that are *not* manipulative and that, instead, engage psychological capacities that underlie rational agency.

## 1. Nudges

Let's consider another illustrative example of nudging. In the U.S. and many other countries obesity has become an epidemic with staggering costs on well-being and public health care systems. In response to this problem coercive measures are sometimes taken. Under the leadership of Michael Bloomberg, for example, New York City attempted (though ultimately failed) to ban soda drinks over 16 ounces in size. The stakes here are not high, admittedly, but many critics objected to what they saw as an unwelcome intrusion by the government into matters of personal choice.

Thaler and Sunstein recommend nudging people instead of imposing legal constraints on them. Cafeterias, instead of eliminating any products, might simply reorganize the way in which food is displayed (2008: 1-3). Vegetables could be placed at eye-level, say, while french fries could be placed below or above eye-level. The likely result would be a significant improvement in immediate food choices and also long-term eating habits. And yet, once again, freedom of choice is preserved. Unhealthy food remains an option for those who prefer it.

Thaler and Sunstein's work is full of striking examples, and we'll consider more examples as needed. However, a preliminary difficulty for the authors is that, for all their fascinating illustrations, they do not provide a sufficiently precise, general definition of nudges. Thaler and Sunstein do say, quite clearly, that nudges do not coercively limit a person's options and that they do not introduce new incentives or disincentives. Avoiding the effects of a nudge, they say, should be relatively costless. Furthermore, the authors note that nudges may be either paternalistic or altruistic. That is, nudges may influence people's behavior for the sake of their own interests or for the sake of other people's interests. Re-organizing food displays in cafeterias is paternalistic. Switching to opt-out organ donation is altruistic.

Despite these clarifications, Thaler and Sunstein's working definition of a nudge remains far too broad. To see this, consider other ways of influencing behavior that, intuitively, do not count as nudges. Simple warnings can be used to influence behavior. So can rational arguments. Both of these methods avoid coercion and incentives, and both may be used for either paternalistic or altruistic ends. If the concept of a nudge subsumes

simple warnings and rational arguments, however, it is not the unified category of moral shortcuts that it purports to be.

What's missing? Indeed, something else unites Thaler and Sunstein's striking examples. Nearly all of their proposed nudges exploit so-called "heuristics and biases" in judgment and decision-making (Bovens 2008; Hausman and Welch 2010; cf. Wilkinson 2013: 343). Heuristics and biases are useful shortcuts for reasoning. They work reliably in familiar environments, and, indeed, on the whole are probably indispensable tools in judgment and decision-making (see, e.g., Gigerenzer and Goldstein 1996). However, in novel or unusual contexts, heuristics and biases can lead one astray, resulting in judgments and decisions that frustrate an agent's interests, even by her own lights.

Thaler and Sunstein advocate what they call "libertarian paternalism" (cf. Kelly 2016). A critical source of support for libertarian paternalism is the existence of heuristics and biases. John Stuart Mill (1859) famously argued that paternalism is morally problematic—even on consequentialist grounds—because a person herself is usually in a better position than others to know what is in her own interests. However, the distorting influence of heuristics and biases contradicts Mill's epistemic assumption, according to Thaler and Sunstein. People make systematic errors about how to satisfy their own interests, and this is why nudges are important tools for helping people to lead better lives (see esp. Sunstein 2012: 6-11). To nudge is to exploit heuristics and biases in judgment and decision-making—except in ways that lead to positive outcomes rather than negative outcomes. To see this, and to arrive at an adequate definition of nudges, let's look again at some of Thaler and Sunstein's illustrative examples.

Thaler and Sunstein argue that people have a general bias to stick with the status quo. This bias explains why switching to opt-out organ donation functions as a nudge. The "status quo bias" also underlies several other key examples of nudges. Most people who enter into retirement savings plans subsequently make no changes to their asset allocations (Thaler and Sunstein 2008: 34). Many elderly patients fail to enroll in prescription drug plans provided by Medicare, and they are reluctant to switch to plans that serve them better (2008: 159-62). In both cases, Thaler and Sunstein suggest that choice architects can use expert analysis to enroll people in plans that are likely to best suit their needs—as a default—while also giving them the option to switch easily if they so choose. These nudges capitalize on the status quo bias while also, once again, preserving freedom of choice.

The status quo bias is a motivational bias. But people also possess various cognitive biases that disrupt their ability to process information correctly. Attentional biases underlie the other case with which we began: reorganizing cafeteria food to present healthy food at eye-level. Another cognitive bias that Thaler and Sunstein discuss involves disproportionate aversion to loss. One is likely to care more about potential losses than potential gains even when the values at play are identical, a clear violation of cost-benefit reasoning with no clear rationale. Thus, Thaler and Sunstein suggest that doctors and oth-

er health care providers can nudge patients into making better health care choices by presenting important medical information in terms of the costs of forgoing treatment rather than gains of pursuing treatment (2008: 36).

As a final illustration of the role that heuristics and biases play in nudges, let's turn to Thaler and Sunstein's discussion of social pressures that bias people to conform to others' behavior. The authors suggest that circulating facts about the typical behavior of others can increase rates of tax compliance (2008: 66), discourage harmful treatment of the environment (66-7), decrease binge drinking among teens (67-8), and lower wasteful and unnecessary energy consumption (68-9). Many people have inaccurate information about the frequency with which others take part in these activities. In virtue of an underlying bias to conform, imparting more accurate information can nudge people's behavior so that it is closer to ideal levels—ideal in the sense that it brings about positive outcomes either for the targeted agents themselves or for others.

In light of all of this, the following definition presents itself: a nudge is a way of using choice architecture to alter people's behavior (1) without coercion or incentives, (2) either paternalistically or altruistically, and (3) *via common heuristics and biases*. This expanded definition has two main virtues. First, it provides a unifying explanation of Thaler and Sunstein's examples of nudges. Second, it rules out methods for altering behavior that do not seem to fall within the authors' intended scope, i.e., those that do not operate via common heuristics and biases, like simple warnings and rational arguments.

As we'll see next, however, this definition brings into focus a moral challenge to nudging. Before we shift from conceptual inquiry to normative criticism, however, it is worth noting that Thaler and Sunstein explicitly resist an expanded definition of nudges along these lines (see esp. Sunstein 2012: ch. 5). We'll return to this conceptual issue later on in the essay, and I'll argue that the normative criticism I raise does not stand or fall with it. For now, however, it suffices to say that the centrality of heuristics and biases to the authors' many concrete examples belies their resistance to explicitly defining nudges in terms of them.

## 2. Virtues & Vices

According to Thaler and Sunstein, nudges have a flood of virtues and virtually zero vices. Nudges allow choice architects to influence people in ways that are good for them and for others. They can do so at little financial cost—or anyway are often less costly than alternatives. Moreover, Thaler and Sunstein argue that nudges preserve freedom of choice and avoid morally problematic coercion. Nudging does not force people to take any particular course of action, and so leaves it open for them to go their own way. People are nudged rather than shoved.

For Thaler and Sunstein, nudges are a tool of libertarian paternalism. That is, nudges appeal to the libertarian desire to preserve freedom of choice and also to the paternalistic desire to influence people in ways that leave them better off. The combination of libertar-

ianism and paternalism may seem paradoxical, but the authors argue persuasively that it is not. When behavioral interventions avoid coercion, paternalism is, it seems, compatible with freedom of choice. On these and other grounds, Thaler and Sunstein suggest that nudges are likely to garner bi-partisan support in the U.S. political system (2008: 13-4). Nudges appeal to Republicans' desire for minimal government interference and balanced budgets, but they also appeal to Democrats' desire for policies that help people live flourishing lives. All of this seems almost too good to be true.

In the previous section, I developed a needed definition of nudges that is absent from Thaler and Sunstein's work, though crucial to make sense of it. The final condition in this definition states that nudges operate by exploiting common heuristics and biases. This condition, however, leads to a moral difficulty. Nudges do not coercively limit a person's choices—people can still opt out of organ donation, still choose unhealthy food in cafeterias, still consume energy at high levels, etc. Nonetheless, nudges violate autonomy. The reason is that because nudges exploit heuristics and biases, they *bypass* psychological capacities that underlie rational agency. Nudges do not constrain people by force or by threats, but they violate autonomy in another way: by *circumventing* rational agency.

As Martin Wilkinson (2013) puts it, nudges are manipulative. To illustrate, notice that nudges take advantage of people in much the same way that commercial advertising often does. Commercial advertising is morally problematic on two counts. First of all, it tends to influence consumers' behavior in ways that do not serve their legitimate interests or anyone else's, and only serve corporate interests of dubious worth. But commercial advertising is also problematic—apart from its negative consequences—because it takes advantage of people by manipulating them. Even well designed nudges, which avoid the first vice of commercial advertising, are guilty of the second.

An analogy between nudges and commercial advertising is suggestive, but why exactly should we think that nudges are manipulative? Of course, nudges manipulate people in the purely descriptive sense that they change their behavior. But in what morally problematic sense are they manipulative? An action is manipulative, in general, if it is a deliberate attempt to change someone's behavior via non-rational, non-autonomous means. For example, suppose you give someone an argument that you hope will change her mind. If you want her to change her mind by appreciating the soundness of the argument, then your action is not manipulative (not in any morally problematic sense). But if you hope she will respond to evocative language in your argument, and thus be moved non-rationally to change her mind, then voicing the very same words is manipulative. Normally, we think that someone is being manipulative when he is influencing another person for his own benefit. But nudging is different: it threatens to be a form of *manipulative paternalism*. Nudges, that is, may bring about positive outcomes, but they do so by using people, by treating them as mere means and not as ends (cf. Hasnas 2016).

What might Thaler and Sunstein say in their defense? Well, first of all, they may accept the criticism and insist anyway that nudging is better than any alternatives—better

than coercion and better than nothing. Nudges are manipulative and therefore do not completely succeed by deontological standards, the response goes, but they are nonetheless critical tools for improving people's lives. This is a fair point. What I have offered, to be sure, is a reason not to reject nudges entirely, but to recognize their ethical limits. And as we weigh different values that bear on governance and management we may find that some nudges, though manipulative, are still worth implementing. To be clear: my conclusion is not that nudges are manipulative and therefore morally impermissible. Rather, I am urging that the manipulative character of nudges is an unacknowledged moral cost that must be counted against them, alongside any benefits. To responsibly assess and implement nudges we must recognize that they are *not* moral loopholes.

Judith Lichtenberg (2016) similarly notes that even if nudges are manipulative, they may be the best option, all things considered. However, she also argues that some nudges are not manipulative. For example, suppose someone uses air freshener to make a room smell nice in an effort to make people more agreeable and, say, more likely to make a purchase that is in their own best interests. Intuitively, of course, using air freshener is morally unobjectionable. However, I think Lichtenberg is wrong to insist that it is not manipulative. Rather, it is simply manipulative to a very minor degree. Indeed, life is full of such relatively innocuous manipulation. In general, perhaps, an action is more or less manipulative depending on the degree to which it leads people away from the actions they would otherwise choose. Perhaps, furthermore, manipulation is more or less morally objectionable depending on the importance of the choice at hand. Still, because nudges circumvent rational agency they are to some degree manipulative.

All sides should agree that even if nudges are manipulative, that is not a decisive strike against them. However, Thaler and Sunstein are also likely to lean on another, less concessive response. There is, they think, a "misconception that it is possible to avoid influencing people's choices. In many situations, some organization or agent *must* make a choice that will affect the behavior of some other people. There is, in those situations, no way of avoiding nudging in some direction" (2008: 10). And so, even though nudges are manipulative, and even though manipulation is morally problematic, one cannot avoid manipulation (see also Lichtenberg 2016). Better, then, to nudge in ways that positively impact people's lives, rather than in ways that are neutral or, indeed, negative.

The limits of this response lie in its explicit qualifications: "in many situations" and "in those situations." If a choice architect designing a service or policy must face the prospect of influencing behavior in one or another way, then, yes, it seems that nudging people in the best ways possible is, well, best. However, some nudges target contexts of choice that are not already structured or that have less structure than the choice architect proposes, while other nudges target contexts of choice that are not already structured by intentional design. Let's take each of these two possibilities in turn.

First, many of Thaler and Sunstein's proposed nudges introduce novel structure on contexts of choice. For example, they argue that social conformity biases alter people's

behavior so that it is in line with statistical norms, and thus that communicating information about the frequency with which other people comply with tax law and consume energy, say, can lead people to increase their rate of tax compliance and decrease their energy consumption, respectively. But here the authors propose *additional* structural influences on choice that are not already present. Of course, the proposed changes in behavior lead to seemingly desirable outcomes, but the means through which the outcomes are effected introduce novel sources of manipulation where none existed before. This sort of manipulation may be worth it, all things considered, but my point here is that it is not unavoidable.

Second, some choice contexts are already structured, but not in virtue of intentional design by choice architects. For example, cafeteria food is inevitably organized in ways that place some products at eye-level and others below or above eye-level. So, attentional biases cannot but lead patrons to choose some products over others. However, the inevitability of influence does not entail that nudging is inevitable. Nudging is the deliberate structuring of choice contexts, and where that structure is not the product of intentional design it does not count as a nudge. There can be no manipulation without intentional action, and therefore in the cases at hand nudging introduces manipulation where none existed before. (See also Kevin Vallier (2016), who argues that nudging is not only an intentional action but also requires certain sorts of beliefs and motives on the part of choice architects.) Strict consequentialists will see here a distinction without a difference: all that matters are what choices people make. According to consequentialists, intended structure that improves people's lives is better than unintended structure that doesn't. But the moral value at stake here—the value of autonomy—is not consequentialist. The worry remains that nudging uses people because it is manipulative.

The challenge I have been pressing against nudges is that they exploit heuristics and biases and therefore that they violate autonomy by circumventing rational agency. For reasons that will soon be evident, I want to discuss another, related challenge, this one practical rather than moral. The challenge is that interventions mediated by heuristics and biases tend to be *unstable*. Once people realize how they are being manipulated, they are likely to make compensatory modifications to their behavior in response. Thus, while nudges can effect valuable gains in the short run, the gains may be lost in the long run.

To illustrate, consider Thaler and Sunstein's (2008: 38-9) discussion of Lake Shore Drive in Chicago. On one stretch of this road is a series of sharp curves. The speed limit is lowered there. But drivers often ignore the posted signs, maintain a high level of speed, and increase the risk of accidents. To address this problem, the city has implemented what Thaler and Sunstein see as a nudge. The white stripes that divide lanes have been painted closer together as the sharp curves approach. Because of perceptual biases, drivers get the (false) sense that their speed is increasing and automatically slow down, thus mitigating the risk of accidents.

So far, so good. However, once drivers realize how their perceptual system is being tricked, they are likely to override their senses and maintain their high level of speed. Once nudged they *bounce back*. Similar possibilities arise for other nudges. Once people recognize that their cafeteria choices are being influenced by attentional biases, they may be disposed to range their eyes more widely as they decide what to eat for lunch. Once people realize that they are being led to conform to others' behavior, they may be disposed to rebel against conformity. The possibility that people may bounce varies from nudge to nudge. For difficult and exhausting decisions, many people may welcome nudges and evince no inclination to bounce. Nonetheless, bounces are often a real practical problem that choice architects must face, on pain of ultimately losing any gains achieved by nudges.

Another moral problem for nudges, though one that I will not linger on, is that even if nudges are good in principle, they can be misused in practice. For example, although government agents can implement nudges that have the power to dramatically improve people's lives, it is not obvious they should be trusted to wield such power. For one thing, we might doubt their competence as choice architects. For another, government decisions are often influenced by corporate interests and lobbying groups. The result is that some nudges may serve the interests of people other than those they are supposed to help. One solution to this problem is to require transparency: proposed nudges should be widely discussed in the public sphere and advertised to those they are intended to affect. A transparency condition on nudges would seem to be a moral imperative. However, transparency leads people to become more aware of the way in which their judgments and decisions are influenced, and therefore makes bounces even more likely, deepening the practical problem of instability.

To nudge is, among other things, to exploit heuristics and biases in judgment and decision-making. A consequent moral difficulty is that nudges are manipulative. A consequent practical difficulty is instability in nudges' effects. Still, as I have noted, advocates of nudges may insist that these challenges, though genuine, are not decisive. Arguably, the moral benefits of nudging outweigh the moral costs of manipulation. Furthermore, the practical obstacle is worth negotiating as best as we can, even if every bounce requires introducing a further nudge to counter it. So, Thaler and Sunstein may insist that nudging is still the best tool in our kit, all things considered. I want to look for another tool. To find one we must dig deeper into human psychology.

## 3. Bumps

Thaler and Sunstein's case for nudges rests on an empirical view about the human mind, and it is now time to examine that view more closely. To set the stage for their view, Thaler and Sunstein lay out and then reject a competing view of human psychology, supposedly common among economists (at least at one time), that categorizes human beings as "homo economicus," i.e., rational agents who make choices in line with their own values

and desires (2008: 6-8). Of course, human beings are not like this, and accumulated research in psychology and behavioral economics over the past few decades shows that we systematically err in our judgment and decision-making. As Thaler and Sunstein put it, we are not "Econs." Rather, on their view, we are, as I shall say, "Dopes."

Fleshing out their view of human psychology, Thaler and Sunstein appeal to a popular and widely discussed theory of the human mind called a dual process model (see Kahneman 2011). According to a dual process model, there are two types of processes that underlie judgment and decision-making. "Type 1" processes are fast, spontaneous, and unconscious. "Type 2 processes" are slow, deliberative, and conscious. Much of our behavior, perhaps even the majority of it, seems to be driven by type 1 processes. Furthermore, Thaler and Sunstein characterize type 1 processes as consisting largely in heuristics and biases. In other words, type 1 processes are Dopey. From these two assumptions, it follows that if we wish to influence behavior, what we should do is target common heuristics and biases.

A growing body of research in psychology and neuroscience, however, complicates the popular interpretation of dual process models upon which Thaler and Sunstein rely. Many psychological processes that are fast, spontaneous, and unconscious embody sophisticated learning algorithms, sensitive to statistical regularities and able to integrate disparate information about risk and reward. Some of the most impressive empirical research has been conducted on type 1 processes that underlie language learning, causal inference, and theory of mind. For example, whereas the Chomskian program in linguistics holds that language must be innate because there is a poverty of stimulus in the language learners' environment, new research suggests that learning mechanisms might underlie natural language acquisition (see, e.g., Perfors et al. 2006, 2011). More interestingly, for present purposes, research on judgment and decision-making also seems to reveal the influence of sophisticated learning mechanisms.

One of the earliest examples of this research comes from studies on the Iowa Gambling Task (Bechara et al. 1994). In this task participants are given the opportunity to select cards from one of two decks. The first deck contains cards that yield big gains but also big losses. The second deck contains cards that yield moderate gains and only small losses. If you want to maximize net gains you should choose cards only from the second deck. Interestingly, participants are able to learn fairly quickly which deck they should draw from, as measured by their behavior, well before they explicitly recognize the difference, much less articulate what explains the difference. Participants report a gut feeling that the second deck is better, and also exhibit galvanic skin responses characteristic of aversion when choosing from the first deck. The learning is fast, spontaneous, and unconscious, and it is correlated with activity in areas of the brain associated with type 1 processing. Indeed, participants with damage to this area of the brain do not learn effectively to maximize gains.

These sorts of findings have led to a reinterpretation of the dual process model of the mind. In short, type 1 process are not uniformly Dopey. Put more abstractly, the distinction between rational and non-rational psychological mechanisms cross cuts the distinction between type 1 and type 2 processes. The empirical findings have also spurred a burgeoning movement in learning theory, both in psychology and in philosophy (for discussion of moral learning, in particular, see Allman and Woodward 2008; Campbell and Kumar 2012; Cushman 2013; Railton 2014; Nichols et al. forthcoming; Kumar 2015; Campbell 2015). The aim of this movement, most plausibly formulated, is not to throw out decades of research on heuristics and biases, but to supplement it with research on rational learning mechanisms. The older heuristics and biases tradition, at least in its monolithic form, was never very plausible in the first place. If we're such Dopes how do we manage to acquire so many rich and sophisticated implicit skills, from playing tennis to writing books? Correction of this popular tradition offers an enriched and more accurate understanding of the human mind. What it suggests, in short, is that human beings are part Dope, but also part Econ (see, e.g., Kahneman and Klein 2009). That aspect of us which is Econ lies not only in our deliberative, type 2 capacities but also in learning mechanisms that fit the general profile for type 1—cognitive processes that are fast, spontaneous, and unconscious.

Now, let's grant that all of this research is of great import for understanding the human mind. What is its significance for public policy and management? Learning theory in psychology suggests a natural alternative to nudges, what I call "bumps." Bumps are ways of altering people's behavior (1) without coercion or incentives, (2) either paternalistically or altruistically, but (3) *via rational learning mechanisms*. Governments and firms might alter the choice architecture surrounding decisions so as to help people learn to make choices that accurately reflect their own interests, by their own lights. Like nudges, bumps do not force people to act in any particular way. But, unlike nudges, bumps operate through rational agency, rather than bypassing it. Thus, they do not face the same moral problem that casts a shadow over nudges. Bumps are not manipulative.

Recall the practical problem for nudges. Once people become aware that their current behavior is shaped by purposeful exploitation of heuristics are biases, they may bounce back to their previous behavior. Nudges threaten to be unstable. Bumps, however, enable people to learn, effecting psychological changes that are stable and therefore less likely to bounce. Though bumps may take longer than nudges to produce changes in behavior, the changes are more likely to last. So, it seems, bumps have the potential to be both morally and practically better than nudges.

How can bumps be implemented in a way that will promote well being? I have offered a general, abstract understanding of bumps along with their psychological basis in leaning theory. But, so far, we lack a concrete understanding. To make headway on this, let's begin with a clue from Thaler and Sunstein. As they observe, nudges are effective under a limited range of conditions. One condition is that people are unable to receive

regular and informative feedback about their choices. Bumps, by contrast, are likely to be effective only when people *can* receive appropriate feedback about their choices. Now, compared with nudges, there have been fewer real life experiments for us to draw on. Of course, we have plenty of experience with learning, and even with learning that leads to expert judgment and decision-making (see Klein et al. 1993 for review). But there is a relative dearth of real life learning experiments in management and public policy construction. And so, we have no analogue to Thaler and Sunstein's rich stable of examples. Still, some initial suggestions and hypotheses are possible.

For some choices that people face, no regular feedback is available. One can't experiment with life insurance policies and learn how they work out. However, consider medical insurance policies. People who are healthy interact infrequently with medical providers. But chronically sick patients regularly confront the consequences of having chosen a particular form of medical insurance. Choice architects might bump people, then, by presenting them with regular and standardized feedback about the financial costs of their own medical insurance plan along with alternative plans. Patients might receive a simple, standardized form with this information every time they receive medical treatment. We can expect this feedback to attune their implicit learning mechanisms to the risks and rewards associated with different plans. The consequent behavioral changes among patients will be mediated by rational agency and, furthermore, they will be more likely to persist.

One of Thaler and Sunstein's examples is the formally similar RECAP program that enables people to learn to make better choices about credit, loans, and a number of other financial transactions. "RECAP" stands for Record, Evaluate, and Compare Alternative Prices. RECAP is applied across many domains, and one such domain is mortgages, an area of the market in which consumers are regularly offered predatory loans, ultimately a leading cause of the 2008 market crash. One proposed solution to this problem is to ban certain types of loans. However, Thaler and Sunstein argue that this would eliminate loan contracts that can be mutually beneficial. Rather, they suggest that mortgage lenders should be required to uniformly report costs and fees in a simple, salient, accessible way. This would enable consumers to learn more easily which mortgages best serve their interests.

The RECAP program seems to be a bump, in my classification scheme. Thaler and Sunstein, however, think of RECAP as a nudge, and they might, in general, classify a bump as a *type* of nudge. More generally, Thaler and Sunstein are disposed to reject the definition of nudges offered earlier, and on this basis reject the very distinction between nudges and bumps. To fixate on this issue is a mistake, however, since it is merely terminological. We may, instead, think of nudges as the superset, consisting of bumps and what we might call "tricks." Bumps exploit rational learning mechanisms, whereas tricks exploit heuristics and biases. The relative advantages of bumps over nudges, as I've defined them, then, could simply be re-described in terms of the relative advantages of bumps over tricks.

Further taxonomic distinctions might prove useful. For example, in addition to bumps and tricks, consider another type of nudge: "fixes." Fixes are alterations to choice architecture that target computational limitations, rather than heuristics and biases or learning mechanisms. Other, cross-cutting distinctions may be worth drawing too. I have already noted that some interventions are paternalistic, some altruistic (see also Kelly 2016). Paternalistic nudges may rely either on the agent's own conception of his or her welfare or a choice architect's conception (Hasnas 2016). The choice architect may be a government institution or a private firm. Ethical generalizations about nudges may be more secure if the object of evaluation is one or another of these subtypes. Accordingly, my focus has been on certain sorts of nudges, i.e., tricks.

Terminological debate about bumps and nudges and the relationship between them is less important than understanding how choice architects can alter behavior for the better without circumventing rational agency. That's the principal, normative upshot of this essay. Even if nudges do not, by definition, exploit heuristics and biases, some nudges do—i.e., tricks—and thus it seems we have reason to explore behavioral interventions that engage rational learning mechanisms instead.

## 4. Further Discussion

I began the essay by offering some conceptual clarity about nudges and their psychological basis in common heuristics and biases. I then argued that nudges are manipulative and potentially unstable. Drawing on a richer view of human psychology than is evident in Thaler and Sunstein's work, I suggested that bumps—which operate via rational learning mechanisms—avoid these vices of nudging. Now that we have a basic grasp of bumps, however, it's time to evaluate them in more detail.

Bumps are not manipulative, unlike nudges, but they also have two other striking advantages. First, they offer a more farsighted solution to social problems. Nudges alter choice architecture without changing people's underlying psychological dispositions. Thus, their positive outcomes may be lost when people move to new choice contexts. Consequently, too, it is more difficult to build upon nudges in ways that lead to further moral progress. Bumps, by contrast, are *transferable* and *scalable*. Changes to underlying psychological dispositions have the ability to transcend local contexts and they can introduce the possibility of additional psychological changes that lead to further moral progress. For example, a bump that induces people to make better choices about medical insurance might lead them to make better financial decisions in other contexts that share the same structure. And additional bumps become possible that lead to further refinement in their choices.

A second moral advantage of bumps is that they are better suited to accommodate a diversity of values. Effective nudges usually lead all members of a targeted population to one particular outcome. But this is undesirable if relevant interests vary. In that case, a one-size-fits-all nudge is not apt. According to Thaler and Sunstein, remember, nudges

help people to satisfy their interests by their own lights. However, Hasnas (2016) argues that proposed nudges frequently require that choice architects make decisions for people about what is in their interests. Bumps, by contrast, provide people with information which can be deployed in different ways depending on their interests *as they see them*. (Though see Lichtenberg (2016), who argues that choosing what sort of information to convey can be as manipulative as nudging.)

I have been advertising the moral and practical virtues of bumps. However, bumps have vices as well, and it is worth our while to examine them. First, bumps may be financially costly. It's relatively easy to enroll people in a medical insurance plan that is likely to serve them best, relatively difficult to provide them with feedback that will allow them to learn for themselves which plan suits them. Second, bumps generally take longer to work. Default enrollment brings immediate benefits; learning takes time to bear fruit. Third, bumps may be unavailable in cases where nudges are available, for various reasons but among them that people do not have any learning opportunities.

So, bumping may be (a) costly, (b) protracted, or (c) unavailable. In each potential case, choice architects must carry out a careful cost-benefit analysis to examine whether bumping is possible and whether it is worth the effort. If I am right the analysis must include the costs of manipulation. But there are no simple, general answers to the question about whether to nudge or whether to bump. Ethical progress sometimes comes from finding answers to important questions, but at other times it consists in reformulating questions in ways that introduce greater complexity, even if it makes finding answers more difficult. Thus, we should ask ourselves the following questions. When are bumps likely to work? Under what conditions are bumps likely to be more effective than nudges? How long do bumps take to effect valuable gains, and when do the gains begin to drop off? I don't have the answers to these questions, but they are worth raising, and in some cases empirical research in learning theory may help us get approximate answers to them (see, e.g., Slemback and Tyran 2002).

Bumps face practical challenges, as we have seen, but let's turn now to a philosophical challenge. Do bumps, according to the definition that I have offered, *genuinely* facilitate autonomy? Let's bring this challenge into focus: we can imagine cases in which a person's learning mechanisms are engaged, but he is being manipulated. For example, much of the learning that leads teachers to become good at their jobs is implicit and unconscious. They try different things, some of which work and some of which don't. Unconscious reinforcement mechanisms lead teachers to pursue effective strategies and abandon ineffective ones. However, imagine a professor whose students decide to toy cruelly with him. (This is adapted from a story, perhaps apocryphal, about the famous behaviorist B. F. Skinner). Whenever the professor stands on the left side of the classroom the students listen attentively. But whenever he stands on the right side of the classroom the students feign boredom. Eventually, after this schedule of reinforcement, the

professor "learns" to stand only on one side of the classroom. This change in behavior is mediated by rational learning mechanisms, but it is a clear case of manipulation.

I have argued that nudges are manipulative because they target heuristics and biases, whereas bumps are not manipulative because they target rational learning mechanisms. However, it now appears that targeting rational learning mechanisms does not guarantee the absence of manipulation. Our professor's change in behavior is mediated by such mechanisms, but it is the result of manipulation. Now, those who are sympathetic to bumps might argue that simple reinforcement mechanisms are, in fact, non-rational. Indeed, some of the most striking results in learning theory reveal the influence of more sophisticated processes, ones that employ complex internal models of the learner's environment. However, some these sophisticated processes are just elaborations of simple reinforcement learning. Moreover, they also have the potential to be exploited in the same way.

I think the problem here is that the professor's genuine and laudable interest in engaging his students is shaping behavior that he wouldn't *want* to be affected by that interest. Let's put this in more general terms. Learning depends on the agent's interests or values. What sort of learning facilitates autonomy? One condition is that the values in question are guiding behavior that the agent wants—or would want, if informed—to be guided by those values. Thus, an agent wants her values for minimizing financial costs to guide her choice of medical insurance plans. But an agent doesn't want his values for being an effective teacher to guide his choice of where to stand in a classroom. That's why the former isn't manipulative, but the latter is.

As I noted, bumps are likely to be useful when people make decisions for which they receive regular feedback. But bumps require that several other conditions obtain as well: the learning mechanism must be sensitive to the agent's values; the choice problem must be of no more than moderate difficulty; the underlying values must achieve a level of determinacy. Empirical research is needed to identify when and where these conditions hold. Moreover, real life experiments in management and public policy are needed to confirm whether bumps can produce concrete results that improve well being. Still, the promise of bumps is sufficiently high that this work seems worth carrying out.

Bumps seem to have a moral advantage over nudges: they are not manipulative. They also have a practical advantage: they do not threaten instability. I do not deny, however, that nudges are important and ineliminable tools in management and public policy. In many cases it is likely that their virtues outweigh their vices. The reservations about nudges that I have expressed in this essay are cautionary notes, not flashing alarms. Nonetheless, because nudges bypass rational agency, and are therefore manipulative, their vices are more severe than advertised. Fortunately, empirical research in learning theory offers a new tool worth testing. Before you try nudging, think about bumping.

## Acknowledgements

For helpful discussion, I am grateful to Sara Aronowitz, Sameer Bajaj, Richmond Campbell, Stewart Cohen, Ronald de Sousa, Judith Lichtenberg, Shaun Nichols, Peter Railton, and Hannah Tierney. Thanks are owed as well to audiences at the University of Arizona, the 2015 Rocky Mountain Ethics Congress, and especially a symposium on the ethics of nudging held at the Georgetown Institute for the Study of Markets and Ethics.

## Bibliography

Abadie, A. & Gay, S. 2004. "The impact of presumed consent legislation on cadaveric organ donation: a cross country study," NBER Working Paper no. W10604.

Allman, J. & Woodward, J. 2008. "What are moral intuitions and why should we care about them? A neurobiological perspective," *Philosophical Issues*, 18: 164-85.

Ariely, D. 2010. *Predictably Irrational: The Hidden Forces that Shape our Decisions* (New York: Harper Perennial).

Bechara, A., Damasio, A., Damasio, H., & Anderson, S. 1994. "Insensitivity to future consequences following damage to human prefrontal cortex," *Cognition*, 50: 7-15.

Bovens, L. 2008. "The ethics of nudge," in T. Grune-Yanoff & O. Hansson, Eds., *Preference Change: Approaches from Philosophy, Economics, and Psychology* (Berlin: Springer): 207-20.

Campbell, R. 2016. "Consistency reasoning in moral learning: its dual normative functions," Under review.

Campbell, R. & Kumar, V. 2012. "Moral reasoning on the ground," *Ethics*, 122: 273-312.

Cushman, F. 2013. "Action, outcome, and value: a dual system framework for morality," *Personality and Social Psychology Review*, 17: 273-92.

Gigerenzer, G. & Goldstein, D. 1996. "Reasoning the fast and frugal way: models of bounded rationality," *Psychological Review*, 103: 650-69.

Hasnas, J. 2016. "Some noodging about nudging: four questions about libertarian paternalism," *Georgetown Journal of Law and Public Policy*.

Hausman, D. & Welch, B. 2010. "To nudge or not to nudge," *Journal of Political Philosophy*, 18 (1): 123-36.

Johnson, E. & Goldstein, D. 2003. "Do defaults save lives?" *Science*, 302: 1338-9.

Kahneman, D. 2011. *Thinking Fast and Slow* (New York: Farrar, Straus and Giroux).

Kahnemen, D. & Klein, G. 2009. "Conditions for intuitive expertise: A failure to disagree," *American Psychologist*, 64: 515-26.

Kelly, J. 2016. "Market failure nudges," *Georgetown Journal of Law and Public Policy.*

Klein, G., Orasanu, J., Calderwood, R. & Zsambok, C. 1995. *Decision Making in Action: Models and Methods* (Ablex).

Kumar, V. 2015. "Moral vindications," Under review.

Kurts, S. & Saks, M. 1996. "The transplant paradox: overwhelming public support for organ donation vs. under-supply of organs: The Iowa organ procurement study," *Journal of Corporation Law*, 21: 767-806.

Lichtenberg, J. 2016. "For your own good: informing, nudging, coercing," *Georgetown Journal of Law and Public Policy*.

Mill, J. S. 1869. *On Liberty* (London: Longman, Roberts & Green).

Nichols, S., Kumar, S., Lopez, T., Ayars, A. & Chan, H. Forthcoming. "Rational learners," *Mind & Language*.

Perfors, A., Tenenbaum, J. & Regier, T. 2006. "Poverty of the stimulus? A rational approach," in *Proceedings of the 28th Annual Conference of the Cognitive Science Society*: 663-8.

Perfors, A., Tenenbaum, J. & Regier, T. 2011. "The learnability of abstract syntactic principles," *Cognition*, 118: 306-38.

Railton, P. 2014. "The affective dog and its rational tale: intuition and attunement," *Ethics*, 124: 813-59.

Sunstein, C. 2012. *Why Nudge?* (Yale University Press).

Thaler, R. & Sunstein, C. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness* (New Haven, CT: Yale University Press).

Vallier, K. 2016. "Can we never not nudge?" *Georgetown Journal of Law and Public Policy*.

Wilkinson, T. M. 2013. "Nudging and manipulation," *Political Studies*, 2013, 61: 341-55.